INK 000044
3/5

DEMANDE DE BREVET EUROPEEN - 1ER DÉPÔT
EUROPEAN PATENT APPLICATION - 1ST FILING

TITRE OFFICIEL OFFICIAL TITLE	:	VIDEO ENCODING METHOD USING A WAVELET DECOMPOSITION
INVENTEUR(S)-INVENTOR(S)	:	FELTS Boris, PESQUET-POPESCU Béatrice
N° ENREGISTREMENT (DEPOT) OFFICIAL FILING NUMBER	:	00400932.0
DATE DE DEPOT-FILING DATE	:	04/04/2000
DEMANDEUR-APPLICANT Secteur-Sector	: :	KONINKLIJKE PHILIPS ELECTRONICS N.V. Laboratoires d'Electronique Philips
MANDATAIRE-PATENT ATTORNEY	:	LANDOUSY Christian et al. Société Civile SPID 156, Bd Haussmann - 75008 PARIS FRANCE Tél : 33 (0)1 40 76 80 30 - FAX 33 (0)1 45 61 05 36 Liste Spéciale INPI 422.5/S008
Référence CIP-FR (SPID) Correspondant à - corresponding to	: :	PHFR000032 EPp - CLA/CLA/LAND - GG ID 300548

The invention relates to an encoding method for the compression of a video sequence divided in groups of frames decomposed by means of a tridimensional (3D) wavelet transform. Said method is based on a hierarchical subband encoding process called "set partitioning in hierarchical trees" (SPIHT) and leading from the original set of picture elements of the video sequence to transform coefficients encoded with a binary format. These coefficients are ordered by means of magnitude tests involving the pixels represented by three ordered lists called list of insignificant sets (LIS), list of insignificant pixels (LIP) and list of significant pixels (LSP).

According to the invention, the spatio-temporal approximation subband resulting from the 3D wavelet transform contains the spatial approximation subbands of the two frames in the temporal approximation subband, indexed by $z = 0$ and $z = 1$, and a specific initialization order is proposed. Moreover, the spatio-temporal orientation trees defining the spatio-temporal relationship on the hierarchical pyramid of the wavelet decomposition are explored from the lowest resolution level to the highest one, while keeping neighbouring pixels together and taking account of the orientation of the details.

«VIDEO ENCODING METHOD USING A WAVELET DECOMPOSITION»

FIELD OF THE INVENTION

The present invention relates to an encoding method for the compression of a video sequence divided in groups of frames decomposed by means of a tridimensional (3D) wavelet transform leading to a given number of successive resolution levels, said method being based on a hierarchical subband encoding process called "set partitioning in hierarchical trees" (SPIHT) and leading from the original set of picture elements of the video sequence to transform coefficients encoded with a binary format, said coefficients being ordered by means of magnitude tests involving the pixels represented by three ordered lists called list of insignificant sets (LIS), list of insignificant pixels (LIP) and list of significant pixels (LSP), said tests being carried out in order to divide said original set of picture elements into partitioning subsets according to a division process that continues until each significant coefficient is encoded within said binary representation.

BACKGROUND OF THE INVENTION

Classical video compression schemes may be considered as comprising four main modules : motion estimation and compensation, transformation in coefficients (for instance, discrete cosine transform or wavelet decomposition), quantification and encoding of the coefficients, and entropy coding. When a video encoder has moreover to be scalable, this means that it must be able to encode images from low to high bit rates, increasing the quality of the video with the rate. By naturally providing a hierarchical representation of images, a transform by means of a wavelet decomposition appears to be more adapted to scalable schemes than the conventional discrete cosine transform (DCT).

A wavelet decomposition allows an original input signal to be described by a set of subband signals. Each subband represents in fact the original signal at a given resolution level and in a particular frequency range. This decomposition into uncorrelated subbands is generally implemented by means of a set of monodimensional filter banks applied first to the lines of the current image and then to the columns of the resulting filtered image.

An example of such an implementation is described in "Displacements in wavelet decomposition of images", by S. S. Goh, Signal Processing, vol. 44, n° 1, June 1995, pp.27-38. Practically two filters – a low-pass one and a high-pass one – are used to separate low and high frequencies of the image. This operation is first carried out on the lines and followed by a sub-sampling operation, by a factor of 2. It is then carried out on the columns of the sub-sampled image, and the resulting image is also down-sampled by 2. Four images, four times smaller than the original

one, are thus obtained : a low-frequency sub-image (or "smoothed image"), which includes the major part of the initial content of the concerned original image and therefore represents an approximation of said image, and three high-frequency sub-images, which contain only horizontal, vertical and diagonal details of said original image. This decomposition process continues until it is clear that there is no more useful information to be derived from the last smoothed image.

A technique rather computationally simple for image compression, using a wavelet decomposition, is described in "A new, fast, and efficient image codec based on set partitioning in hierarchical trees (= SPIHT)", by A. Said and W.A. Pearlman, IEEE Transactions on Circuits and Systems for Video Technology, vol.6, n°3, June 1996, pp.243-250. As explained in said document, the original image is supposed to be defined by a set of pixel values p(i,j), where i and j are the pixel coordinates, and to be coded by a hierarchical subband transformation, represented by the following formula (1) :

$$c(i,j) = \Omega\,(p(i,j)) \qquad (1)$$

where Ω represents the transformation and each element c(i,j) is called "transform coefficient for the pixel coordinates (i,j)". The major objective is then to select the most important information to be transmitted first, which leads to order these transform coefficients according to their magnitude (coefficients with larger magnitude have a larger content of information and should be transmitted first, or at least their most significant bits). If the ordering information is explicitly transmitted to the decoder, images with a rather good quality can be recovered as soon as a relatively small fraction of the pixel coordinates are transmitted.

If the ordering information is not explicitly transmitted, it is then supposed that the execution path of the coding algorithm is defined by the results of comparisons on its branching points, and that the decoder, having the same sorting algorithm, can duplicate this encoder's execution path if it receives the results of the magnitude comparisons. The ordering information can then be recovered from the execution path.

One important fact in said sorting algorithm is that it is not necessary to sort all coefficients, but only the coefficients such that $2^n \le |c_{i,j}| < 2^{n+1}$, with n decremented in each pass. Given n, if $|c_{i,j}| \ge 2^n$ (n being called the level of significance), it is said that a coefficient is significant ; otherwise it is called insignificant. The sorting algorithm divides the set of pixels into partitioning subsets T_m and performs the magnitude test (2) :

$$\max_{(i,j)\in T_m} \left\{ \left| c_{i,j} \right| \right\} \ge 2^n \quad ? \qquad (2)$$

If the decoder receives a "no" (the whole concerned subset is insignificant), then it knows that all coefficients in this subset T_m are insignificant. If the answer is "yes" (the subset is

significant), then a predetermined rule shared by the encoder and the decoder is used to partition T_m into new subsets $T_{m,\ell}$ and the significance test is further applied to these new subsets. This set division process continues until the magnitude test is done to all single coordinate significant subsets in order to identify each significant coefficient.

To reduce the number of transmitted magnitude comparisons (i.e. of message bits), one may define a set partitioning rule that uses an expected ordering in the hierarchy defined by the subband pyramid. The objective is to create new partitions such that subsets expected to be insignificant contain a large number of elements, and subsets expected to be significant contain only one element. To make clear the relationship between magnitude comparisons and message bits, the following function is used :

$$S_n(T) = \begin{cases} 1, & \max\limits_{(i,j)\in T} \left\{ \left| c_{i,j} \right| \right\} \geq 2^n, \\ 0, & \text{otherwise}, \end{cases} \qquad (3)$$

to indicate the significance of a subset of coordinates T.

Furthermore, it has been observed that there is a spatial self-similarity between subbands, and the coefficients are expected to be better magnitude-ordered if one moves downward in the pyramid following the same spatial orientation. For instance, if low-activity areas are expected to be identified in the highest levels of the pyramid, then they are replicated in the lower levels at the same spatial locations, but with a higher resolution . A tree structure, called spatial orientation tree, naturally defines the spatial relationship on the hierarchical pyramid of the wavelet decomposition. Fig.1 shows how the spatial orientation tree is defined in a pyramid constructed with recursive four-subband splitting. Each node of the tree corresponds to the pixels of the same spatial orientation in the way that each node has either no offspring (the leaves) or four offspring, which always form a group of 2 x 2 adjacent pixels. In Fig.1, the arrows are oriented from the parent node to its four offspring. The pixels in the highest level of the pyramid are the tree roots and are also grouped in 2 x 2 adjacent pixels. However, their offspring branching rule is different, and in each group, one of them (indicated by the star in Fig.1) has no descendant.

The following sets of coordinates are used to present this coding method, (i,j) representing the location of the coefficient) :

. O(i,j) : set of coordinates of all offspring of node (i,j) ;

. D(i,j) : set of coordinates of all descendants of the node (i,j) ;

. H : set of coordinates of all spatial orientation tree roots (nodes in the highest pyramid level) ;

. L(i,j) = D(i,j) - 0(i,j).

As it has been observed that the order in which the subsets are tested for significance is important, in a practical implementation the significance information is stored in three ordered lists, called list of insignificant sets (LIS), list of insignificant pixels (LIP), and list of significant pixels (LSP). In all these lists, each entry is identified by coordinates (i,j), which in the LIP and LSP represent individual pixels, and in the LIS represent either the set D(i,j) or L(i,j) (to differentiate between them, a LIS entry may be said of type A if it represents D(i,j), and of type B if it represents L(i,j)). The SPIHT algorithm is in fact based on the manipulation of the three lists LIS, LIP and LSP.

For the entropy coding module, the arithmetic coding is a widespread technique which is more effective in video compression than the Huffmann encoding owing to the following reasons : the obtained codelength is very close to the optimal length, the method particularly suits adaptive models (the statistics of the source are estimated on the fly), and it can be split into two independent modules (the modeling one and the coding one). The following description relates mainly to modeling, which involves the determination of certain source-string events and their context, and the way to estimate their related statistics. The context is intended to capture the redundancies of the entire set of source strings under consideration.

In the original video sequence, the value of a pixel indeed depends on those of the pixels surrounding it. After the wavelet decomposition, the same property of "geographic" interdependency holds in each spatio-temporal subband. If the coefficients are sent in an order that preserves these dependencies, it is possible to take advantage of the "geographic" information in the framework of universal coding of bounded memory tree sources, as described for instance in the document "A universal finite memory source", by M.J. Weinberger and al., IEEE Transactions on Information Theory, vol.41, n°3, May 1995, pp.643-652. A finite memory tree source has the property that the next symbol probabilities depend on the actual values of the most recent symbols. Binary sequential universal source coding procedures for finite memory tree sources often make use of context tree which contains for each string (context) the number of occurrences of zeros and ones given the considered context. This tree allows to estimate the probability of a symbol, given the d previous bits :

$\hat{P}(X_n | x_{n-1} \ldots x_{n-d})$, where x_n is the value of the examined bit and $x_{n-1} \ldots x_{n-d}$ represents the context, i.e. the previous sequence of d bits. This estimation turns out to be a difficult task when the number of conditioning events increases because of the context dilution problem or the model cost. One way to solve this problem is the context-tree weighting method, detailed in "The context-tree weighting method : basic properties", by F.M.J. Willems and al., IEEE Transactions on Information Theory, vol.41, n°3, May 1995, pp.653-664.

The principle of this method is to estimate weighted probabilities using the most efficient context for the examined bit. Indeed, sometimes it can be better to use shorter contexts to encode a bit (if the last bits of the context have no influence on the current bit, they might not be taken into account). This technique reduces the length of the final code. The determination of efficient models and contexts is therefore a crucial stage in arithmetic encoding.

Unfortunately, the SPIHT algorithm, which exploits the redundancy between the subbands, "destroys" the dependencies between neighbouring pixels inside each subband. The 2D SPIHT algorithm is based on a key concept : the prediction of the absence of significant information across scales of the wavelet decomposition by exploiting self-similarity inherent in natural images. This means that if a coefficient is insignificant at the lowest scale of wavelet decomposition, the coefficients corresponding to the same area at the other scales have great chances to be insignificant too. Basically, the SPIHT algorithm consists in comparing a set of pixels corresponding to the same image area at different resolutions to the value previously called "level of significance".

The 3D SPIHT algorithm does not differ greatly from the previous one. A 3D-wavelet decomposition is performed on a group of frames (GOF). Following the temporal direction, a motion compensation and a temporal filtering are realized. Instead of spatial sets (2D), one has 3D spatio-temporal sets, and trees of coefficients having the same spatio-temporal orientation and being related by parent-offspring relationships can be also defined. These links are illustrated in the 3D case in Fig.2. The roots of the trees are formed with the pixels of the approximation subband at the lowest resolution ("root" subband). In the 3D SPIHT algorithm, in all the subbands but the leaves, each pixel has 8 offspring pixels, and mutually, each pixel has only one parent. There is one exception at this rule : in the root case, one pixel out of 8 has no offspring.

As in the 2D case, a spatio-temporal orientation tree naturally defines the spatio-temporal relationship on the hierarchical wavelet decomposition, and the following sets of coordinates are used :

. O(x,y,z chroma) : set of coordinates of all offspring of node (x,y,z chroma) ;

. D(x,y,z chroma) : set of coordinates of all descendants of the node (x,y,z chroma) ;

. H(x,y,z chroma) : set of coordinates of all spatio-temporal orientation tree roots (nodes in the highest pyramid level) ;

. L(x,y,z, chroma) = D(x,y,z, chroma) - O(x,y,z, chroma) ;

where (x,y,z) represents the location of the coefficient and "chroma" stands for Y, U or V. There are three ordered lists : LIS (list of insignificant sets), LIP (list of insignificant pixels), LSP (list of significant pixels). In all these lists, each entry is identified by a coordinate (x,y,z, chroma), which in the LIP and LSP represents individual pixels, and in the LIS represents one of D(x,y,z, chroma) or L(x,y,z, chroma) sets. To differentiate between them, the LIS entry is of type A if it represents D(x,y,z, chroma), and of type B if

it represents L(x,y,z, chroma). As previously in the 2D case, the algorithm 3D SPIHT is based on the manipulation of these three lists LIS, LIP and LSP.

However, the manipulation of the lists LIS, LIP, LSP, conducted by a set of logical conditions, makes the order of pixel scanning hardly predictable. The pixels belonging to the same 3D offspring tree but from different spatio-temporal subbands are encoded and put one after the other in the lists, which has for effect to mix the pixels of foreign subbands. Thus, the geographic interdependencies between pixels of the same subband are lost. Moreover, since the spatio-temporal subbands result from temporal or spatial filtering, the frames are filtered along privileged axes that give the orientation of the details. This orientation dependency is lost when the SPIHT algorithm is applied because the scanning does not respect the geographic order.

SUMMARY OF THE INVENTION

It is therefore the object of the present invention to improve the scanning order in the SPITH algorithm in order to reestablish the relations of neighbourhood between pixels of the same subband.

To this end, the invention relates to an encoding method such as described in the introductive part of the description and which is moreover characterized in that :

(A) the spatio-temporal approximation subband resulting from the 3D wavelet transform contains the spatial approximation subbands of the two frames in the temporal approximation subband, indexed by z = 0 and z = 1, and, each pixel having coordinates (x,y) varying from 0 to size_x and from 0 to size_y respectively, said list LIS is then initialized with the coefficients of said spatio-temporal approximation subband, excepting the coefficient having the coordinates of the form z=0 (mod 2), x=0 (mod 2) and y=0 (mod 2), said proposed initialization order being the following:

(a) put in the list all the pixels that verify x = 0 (mod.2) and y = 0 (mod.2) and z = 1, for the luminance component Y and then for the chrominance components U and V ;

(b) put in the list all the pixels that verify x = 1 (mod.2) and y = 0 (mod.2) and z = 0, for Y and then for U and V ;

(c) put in the list all the pixels that verify x = 1 (mod.2) and y = 1 (mod.2) and z = 0, for Y and then for U and V ;

(d) put in the list all the pixels that verify x = 0 (mod.2) and y = 1 (mod.2) and z = 0, for Y and then for U and V ;

(B) the spatio-temporal orientation trees defining the spatio-temporal relationship on the hierarchical pyramid of the wavelet decomposition are explored from the lowest resolution level to the highest one, while keeping neighbouring pixels together and taking account of the orientation of the details, said exploration being implemented thanks to a scanning order of the offspring coefficients that is shown in figures 7 to 10.

The initialization of the LIS plays an important role in the progress of the algorithm. A special organization of this list, a particular scan of offspring coefficients and a slight modification of the original algorithm allow to explore the trees in depth while keeping neighbouring pixels together and taking account of the orientation of the details.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will now be described in a more detailed manner, with reference to the accompanying drawings in which :

- Fig.1 illustrates examples of parent-offspring dependencies in the spatial-orientation tree in the 2D case ;
- Fig.2 similarly shows examples of parent-offspring dependencies in the spatio-temporal orientation tree (3D SPIHT) ;
- Fig.3 shows the proposed scanning order of the root subband coefficients having offspring in the horizontal detail subbands ;
- Fig. 4 shows the proposed scanning order of the root subband coefficients having offspring in the diagonal detail subbands ;
- Fig. 5 shows the proposed scanning order of the root subband coefficients having offspring in the vertical detail subbands ;
- Fig.6 illustrates the organized and oriented scan in the detail subbands ;
- Fig.7 describes the proposed scanning order for a group of 4 offspring and the passage from one group to the next one in the horizontal direction (subbands with details having an horizontal or diagonal direction) ;
- Fig.8 describes the proposed scanning order for a group of 4 offspring and the passage from one group to the next one in the vertical direction (subbands with details having a vertical direction) ;
- Fig.9 shows the order of scanning for the coefficients of the lowest resolution subbands ;
- Fig.10 illustrates the scanning order over two resolution levels for subbands with horizontal orientation of details (a special attention is given to passages from one group of pixels to the other, by respecting the proximity of pixels).

DETAILED DESCRIPTION OF THE INVENTION

It has been seen that a main challenge in the efficient insertion of the arithmetic coding into the SPIHT algorithm is to keep the geographic neighbouring in the contexts. The initial organization of the LIS and a particular order of reading the offspring will allow to re-establish partially a geographic scan of the coefficients, as explained first for the 2D SPIHT algorithm restricted to luminance coefficients, and then, as an extension, in the case of the 3D SPIHT algorithm with chrominance components.

The 2D SPIHT algorithm scans the pixels of all the spatial subbands using the parent-offspring dependencies, starting with the coefficients of the root subband that are originally stored in the LIP (and a part of them also in the LIS). By ordering differently the roots of the spatial offspring trees, it is possible to re-establish a coherent order for the examination of the higher subbands one after another, and even to take into account the spatial orientation of the details (the orientation of the details may be indeed better exploited when considering a privileged direction of scanning).

It is therefore proposed to initialize the LIS with the coefficients of the approximation subband (a pixel has coordinates (x,y) with x varying from 0 to size_x and y varying from 0 to size_y) using the following scanning :

(a) put in the list all the pixels (x,y) that verify x = 1 (mod.2) and y = 0 (mod.2), by horizontally scanning the subband (this first case corresponds to Fig.3) ;

(b) put in the list all the pixels (x,y) that verify x = 1 (mod.2) and y = 1 (mod.2), by horizontally scanning the subband (this second case corresponds to Fig.4) ;

(c) put in the list all the pixels (x,y) that verify x = 0 (mod.2) and y = 1 (mod.2), by vertically scanning the subband, which corresponds to the third case of Fig.5 (the pixels (x,y) that verify x = 0 (mod.2) and y = 0 (mod.2) are not inserted in the LIS).

Thanks to this organization of the LIS, the 2D SPIHT algorithm scans the subbands following a prescribed order of examination of the details : the subbands containing the horizontal details are first read, then the subbands containing the diagonal details (for which the order is not so important), and finally the subbands containing the vertical details, from the lowest resolution to the highest one as illustrated in Fig.6.

The invention may be extended to the tridimensional case. This 3D extension is done without any particular initialization on the temporal axis. The temporal approximation subband has two frames, indexed by z = 0 and z = 1, and the proposed initialization order is the following :

(a) put in the list all the pixels that verify x = 0 (mod.2) and y = 0 (mod.2) and z = 1, for the luminance component Y and then for the chrominance components U and V ;

(b) put in the list all the pixels that verify x = 1 (mod.2) and y = 0 (mod.2) and y = 0 (mod.2) and z = 0, for Y and then for U and V ;

(c) put in the list all the pixels that verify x = 1 (mod.2) and y = 1 (mod.2) and z = 0, for Y and then for U and V ;

(d) put in the list all the pixels that verify x = 0 (mod.2) and y = 1 (mod.2) and z = 0, for Y and then for U and V.

The scanning order is vertical in the case (d) and horizontal in the other ones.

The second main aspect of the method consists in a different order of examination of the offspring coefficients. The general rule is that the order of scanning follows the orientation of the details in each subband. This increases the probability of

having large running of ones or zeros, which can be easily compressed by the arithmetic encoder. At each resolution level, a group of 4 offspring coefficients is scanned as depicted in Fig.7 for the horizontal and diagonal detail subbands and as depicted in Fig.8 for vertical detail subbands. An example of scanning at the lowest resolution level is illustrated in Fig.9. This figure describes the scanning order at the pixel level. Pixels are scanned by groups of four. The passage from one group to the other is illustrated in Fig.6 for this passage, the orientation of the details in each subband is followed again (see points 1, 2 and 3 in the figure).

For the finer resolution levels, the scanning order respects the "geographic" proximity, that is, as more as possible, no jump from one line to the other is authorized. Instead, the scanning order proposed in Fig.10 is implemented. The scanning order for groups of four pixels is the same as before, the passage from one group to the other is illustrated in Fig.10 and, at the group level, in Fig.6 (point 4, 5, and 6).

CLAIM :

1. An encoding method for the compression of a video sequence divided in groups of frames decomposed by means of a tridimensional (3D) wavelet transform leading to a given number of successive resolution levels, said method being based on a hierarchical subband encoding process called "set partitioning in hierarchical trees" (SPIHT) and leading from the original set of picture elements of the video sequence to transform coefficients encoded with a binary format, said coefficients being ordered by means of magnitude tests involving the pixels represented by three ordered lists called list of insignificant sets (LIS), list of insignificant pixels (LIP) and list of significant pixels (LSP), said tests being carried out in order to divide said original set of picture elements into partitioning subsets according to a division process that continues until each significant coefficient is encoded within said binary representation, said method being further characterized in that :

(A) the spatio-temporal approximation subband resulting from the 3D wavelet transform contains the spatial approximation subbands of the two frames in the temporal approximation subband, indexed by z = 0 and z = 1, and, each pixel having coordinates (x,y) varying from 0 to size_x and from 0 to size_y respectively, said list LIS is then initialized with the coefficients of said spatio-temporal approximation subband, excepting the coefficient having the coordinates of the form z=0 (mod 2), x=0 (mod 2) and y=0 (mod 2), said proposed initialization order being the following:

(a) put in the list all the pixels that verify x = 0 (mod.2) and y = 0 (mod.2) and z = 1, for the luminance component Y and then for the chrominance components U and V ;

(b) put in the list all the pixels that verify x = 1 (mod.2) and y = 0 (mod.2) and z = 0, for Y and then for U and V ;

(c) put in the list all the pixels that verify x = 1 (mod.2) and y = 1 (mod.2) and z = 0, for Y and then for U and V ;

(d) put in the list all the pixels that verify x = 0 (mod.2) and y = 1 (mod.2) and z = 0, for Y and then for U and V ;

(B) the spatio-temporal orientation trees defining the spatio-temporal relationship on the hierarchical pyramid of the wavelet decomposition are explored from the lowest resolution level to the highest one, while keeping neighbouring pixels together and taking account of the orientation of the details, said exploration being implemented thanks to a scanning order of the offspring coefficients that is shown in figures 7 to 10.

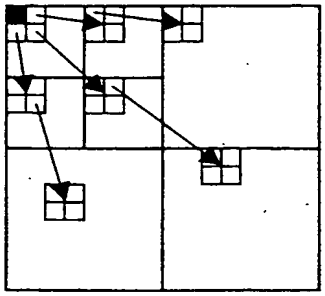

FIG.1

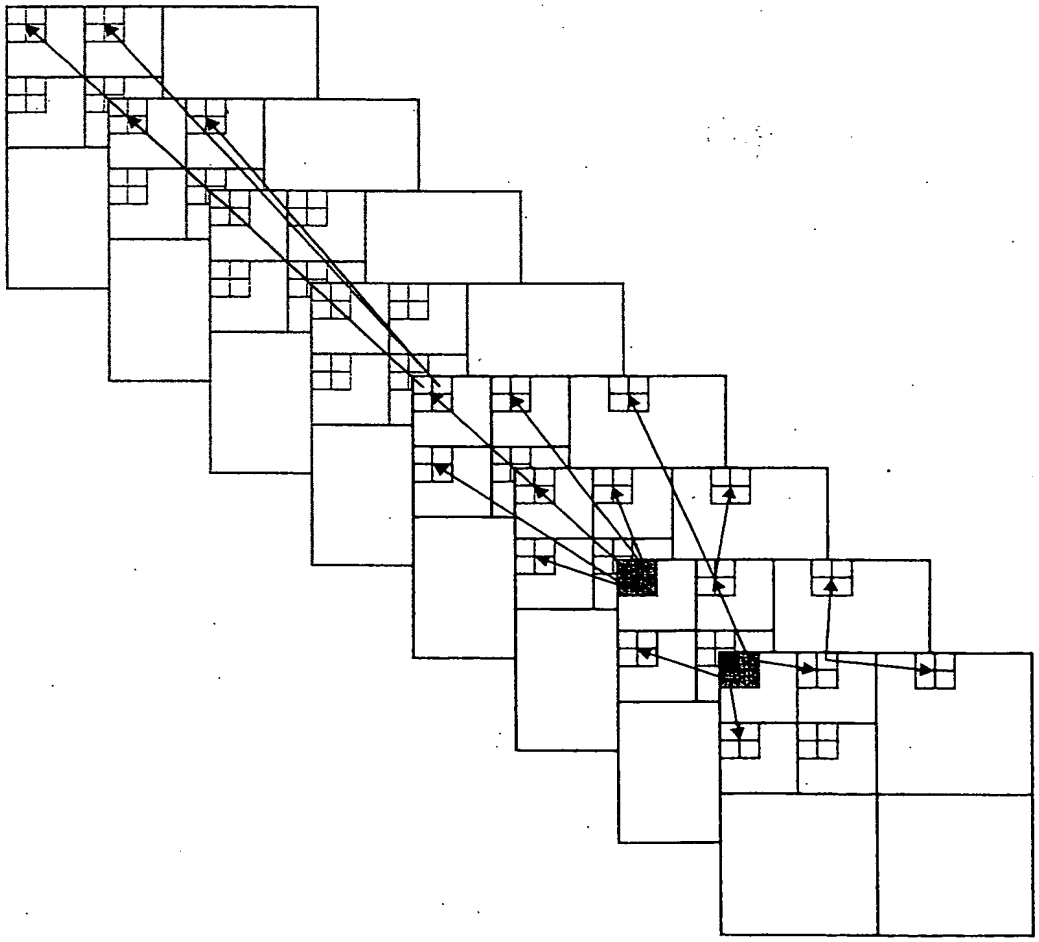

FIG.2

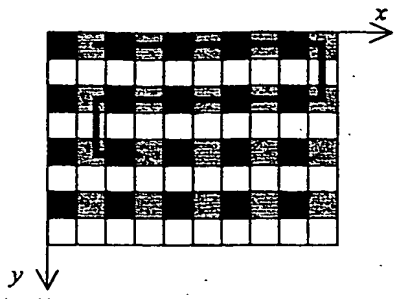

FIG.3

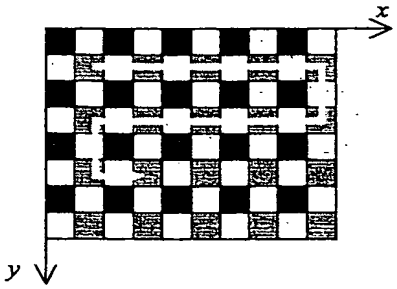

FIG.4

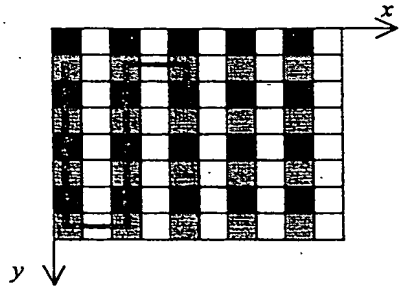

FIG.5

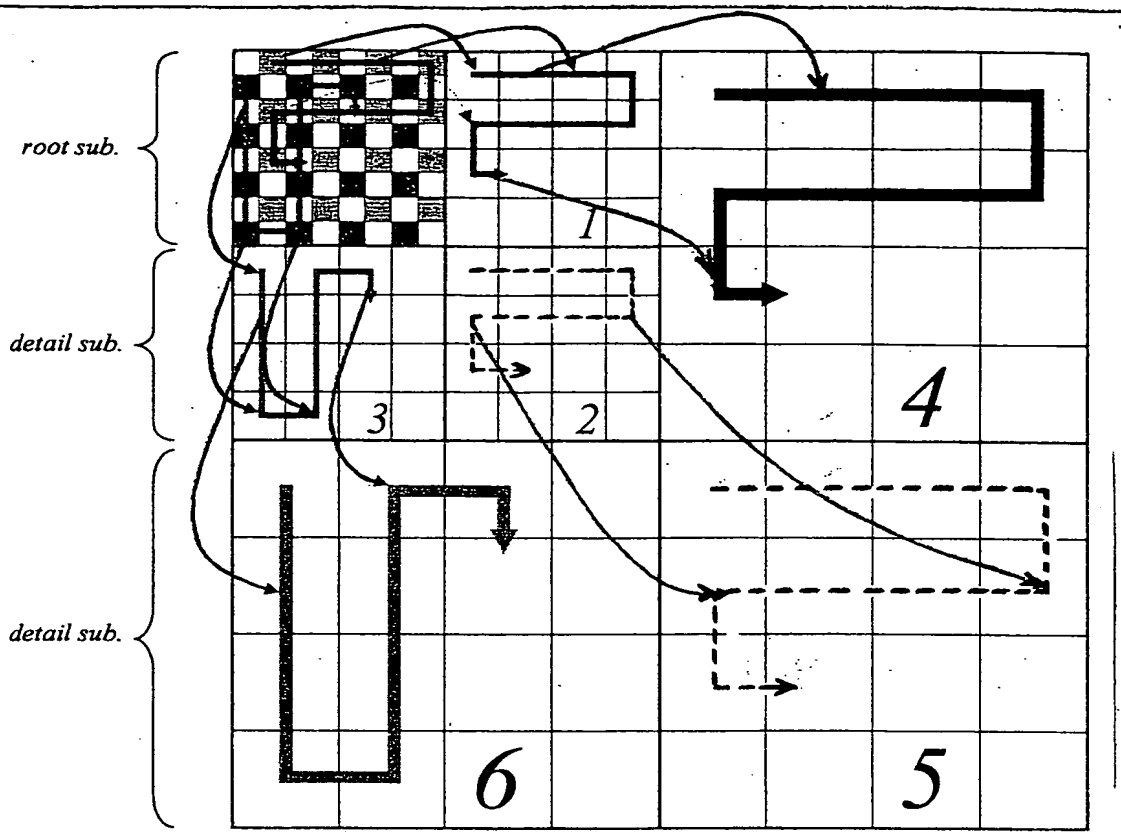

FIG. 6

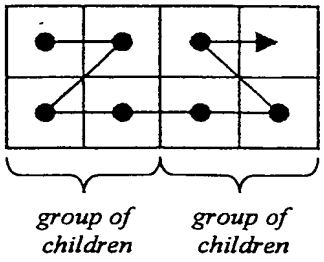

FIG. 7

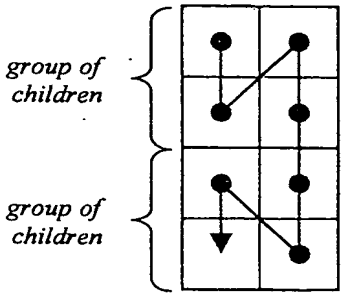

FIG. 8

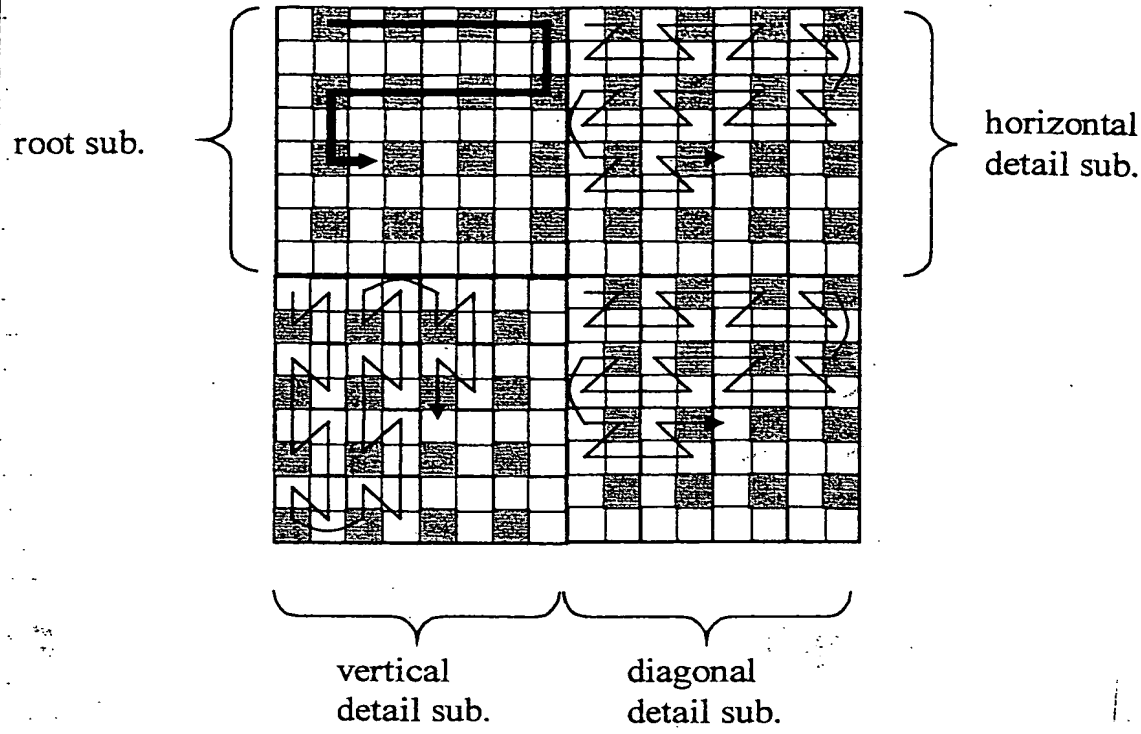
root sub.
horizontal detail sub.
vertical detail sub.
diagonal detail sub.

FIG.9

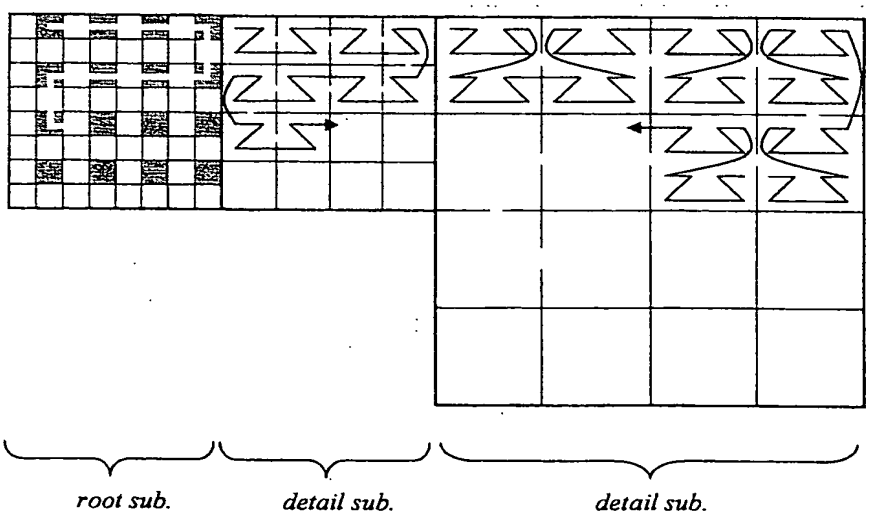
root sub.
detail sub.
detail sub.

FIG.10